

ON CALCULATING THE SQUARE ROOT IN ANCIENT CHINESE MATHEMATICS: A STRANGE METHOD IN SUÀN SHÙ SHŪ

E.F. Scriptor

ABSTRACT

An ancient Chinese collection of problems contains an effective method for calculating square roots. Its logical basis is not explained, and it employs a formula used to solve completely different types of problems. This strange procedure does not appear in later treatises such as the Nine Chapters. This essay explains why this ancient formula succeeds in the goal, not how ancient Chinese mathematicians arrived at it. On this point, one can only speculate.

INTRODUCTION: THE ‘SUÀN SHÙ SHŪ’

Suàn Shù Shū, ‘Writings on reckoning’, is the name of an anonymous ancient Chinese collection of problems of mathematics about seven thousand characters long, contained in 190 bamboo strips.¹ It was found along with other writings in 1983 in a tomb discovered in Mount *Zhangjia*, a locality of the Húběi, a region in central China. The tomb dates back to 186 BC, the beginning of the Western Han Dynasty. “*The Suàn shù shū itself is certainly the oldest Chinese excavated text with*

1 For the translation and interpretation of the text, I relied on the work of C. Cullen, *The Suàn shù shū ‘Writings on reckoning’: A translation of a Chinese mathematical collection...* in *Needham Research Institute Working Papers: 1* ; Cambridge 2004. See [Cullen_NineChapters.pdf](#) .

The italicized sentences in quotation marks " " are taken from the cited article.

substantial mathematical content. Moreover, it is considerably older than any other Chinese mathematical text now extant. Its role in the history of East Asian mathematics is comparable to that of the Ahmose (or Rhind) papyrus in the history of the mathematics of the ancient cultures bordering on the Mediterranean (see Chace (1979) and Gilling (1972)). The importance of the Suàn Shù Shū for the history of world mathematics is therefore indisputable.”

Suàn Shù Shū is a collection of practical problems, accompanied by numerical solutions and related methods. They can be grouped thematically according to different fields of application: Elementary operations with fractions; wastage; sharing, contributions and pricing; changes in rates etc. There are no demonstrations or justifications for the proceedings applied, and the whole treaty is only a collection of examples to illustrate how one must proceed to address and resolve certain problems of an economic, administrative nature. In particular, three problems are solved using the same 'false position' method despite being heterogeneous and it is not easily understood how the procedure adopted can be the same.

ONLY ONE METHOD FOR THREE PROBLEMS

The first of these problems asks how to share out a sum of money.

“In sharing cash, if [each] person [gets] 2 then the surplus is 3. If [each] person [gets] 3 then the deficit is 2. Question: how many persons, and how many cash? Result: 5 persons and 13 cash. Let the excess and the deficit multiply numerators reciprocally to make the dividend. Let the denominators go with one another [in addition] to make the divisor. In each case let the excess or deficit numerator reciprocally multiply the denominator and set each one out separately; reduce the numerator which is greater by the numerator which is lesser; let the remainder be made the divisor; (let the deficit be made the dividend).

The values of two unknowns are asked, the number of people among whom a certain sum must be distributed, and how much this amounts to. It is understood that, once the distribution has been made, nothing remains that has not been distributed. The text offered by the translation is far from clear, so it is best to deduce the solution from the result provided, and then reconstruct the procedure adopted.

The calculated amount of liquidity to be distributed (13 in an unspecified unit of account) is equal to $3 \cdot 3 + 2 \cdot 2$, i.e. the sum of the surplus (3) and deficit (2) each multiplied respectively by the sums that each would receive causing a deficit or a surplus, i.e. in reverse order. Clearly this is a cross product like that of a determinant of 2×2 matrix, but with only positive

quantities. In fact, the expression used is "Let the excess and the deficit *multiply numerators reciprocally* to make the dividend" and then, "let the excess or deficit numerator *reciprocally multiply* the denominator and set each one out separately", which suggests the calculation of a cross product. The terminology used suggests the algebra of fractions, but it seems that the 'denominators' are the surpluses and deficits, and the 'numerators' are the sums possibly distributed if those had the values falsely hypothesized in the procedure adopted (in fact, we speak of the 'false position' method in Western terminology; the Chinese spoke of 'the method of excess and deficit').

The number of people among whom the sum must be distributed (5) is equal to $2 + 3$, that is, the sum of the 'denominators' which is logical, given that by varying by just one unit the sum due to each one goes from a surplus of two units to a deficit of three, with a variation of 5. If the difference had been double, the number of people would also have been double, but it must be noted that this reasoning is correct because the sum distributed to each person varies by only one unit; if it had increased by n units, the sum of surplus and deficit would have had to be divided by the increase in the sum perceived by each person. This is the meaning of the sentence "reduce the major numerator by the minor numerator; the remainder becomes the divisor". In our case, the sum due to each beneficiary varies by only one unit.

To clarify everything, let's write the formulas that express the number of people x and the sum to be distributed y :

$$x = \frac{c+d}{b-a} \quad (\text{F1})$$

$$y = \frac{ad+bc}{b-a} \quad (\text{F2})$$

in which a, b are respectively the sums falsely perceived with $b > a$ ('numerators'), and c, d are the surplus and the deficit ('denominators'). All quantities are positive.

We can obtain the same results by solving two simple linear equations:

$$ax + c = bx - d = y \rightarrow x = \frac{c+d}{b-a}$$

$$y = ax + c = a \frac{c+d}{b-a} + c = \frac{ad+bc}{b-a}$$

At this point, we might ask how ancient Chinese mathematicians had established the general solution of such problems, and the temptation to attribute to them the ability to solve linear systems of two equations in two unknowns would be justified, were it not that the confusing explanation provided to us does not seem to support this hypothesis; but it is true that we are not given theoretical explanations, but only examples of the application of the rules.

Here is another problem, similar to the previous one:

"Now there is a case of common purchase of an ox. If [each group of] 7 families contribute 190, the deficiency is 330; if [each group of] 9 families contribute 270, the excess is 30.

Question: how much are the number of families and the price of the ox? Answer: 126 families; the ox price is 3750.”

Let us assume, as in the previous problem, that x is the number of families, and y is the total sum to be spent. Let's apply F2 to calculate the latter; since each family pays $190/7$ with a deficit of 330 in the first (false) hypothesis, and $270/9$ that is, 30 in the second with a surplus of 30, we must set: $a = 190/7$; $b = 30$; $c = 30$; $d = 330$. For the total sum due, applying F2 we get

$$\frac{\frac{190}{7} \cdot 30 + 30 \cdot 330}{30 - \frac{190}{7}}$$

which is equal to 3750, as expected in the *Suàn Shù Shū*. To calculate the number of families, we apply F1; we get

$$\frac{30+330}{30-190/7} = 126$$

in accordance with the result of the *Suàn Shù Shū*.

THE CALCULATION OF THE SQUARE ROOT

The analysis carried out up to now had the sole aim of obtaining two formulas, of which we are interested in F2 in particular. What is surprising and not entirely explained about it is that, although it is a function of a certain kind of problems, it has a relationship with a procedure useful for calculating square roots. As far as I know, we do not find it in later treatises such as the *Jiuzhang Suanshu* (i.e. *Nine Chapters on the Mathematical Art*), which also takes up many topics covered in the *Suàn Shù Shū*, but calculates the roots in another way.

The problem, with its solution, is as follows.

“Squaring a field. There is a field of one mǔ: how many bù is it square? Reply: it is square 15 bù and 15/31 bù. Method: If it is square 15 bù it is in deficit by 15 bù; if it is square 16 bù there is a remainder of 16 bù. Reply: combine the excess and the deficit to make the divisor; Let the numerator of the deficit multiply the denominator of the excess, and the numerator of the excess multiply the denominator of the deficit; combine to make the dividend. Reverse this like the method for revealing the width.”

'mǔ' and 'bù' are units of area and length, respectively. 1 mǔ equals 240 square bù. So the problem asks to solve the equation $x^2 = 240$. The procedure is as follows: we square 15 and get 225, which is 15 units less than 240; we do the same with 16 and get 256, which exceeds 240 by 16 units. 15 and 16 are the largest of the integer approximations by defect and the smallest of those by excess of the square root of 240, respec-

tively. Let's denote them by a and b respectively. Add the differences of their squares from 240, taken in absolute value, each multiplied by the other approximation, that is, calculate $15 \cdot 16 + 16 \cdot 15 = 480$. Setting $C = 240 - 15^2$ and $D = 16^2 - 240$, the result is divided by $C + D = 31$. In the end, you get

$$\sqrt{240} \approx \frac{480}{31}$$

To check the accuracy of the obtained approximation, let's square it and approximate to ten decimal figures:

$$\left(\frac{480}{31}\right)^2 = 239.7502601$$

which is an excellent approximation of 240, considering that it was obtained immediately with a very simple method, and that it differs from 240 by only one part in a thousand. In fact,

$$\frac{240 - 239.75}{240} = 0.00104166\dots$$

The approximate value of $\sqrt{240}$ reported in *Suàn Shù Shū*, $15 + \frac{15}{31}$, is correct; in fact, $15 + \frac{15}{31} = \frac{480}{31}$.

The proposed solution formula is the ratio of a 'dividend' to a 'divisor' analogous to those of the solution to the first problem examined above: "*Let the excess and deficit multiply each other's numerators to make the dividend. Let the denominators go together [furthermore] to form the divisor*", with the difference that in that case it was necessary to calculate both the amount of a sum of money starting from the 'dividend' and the

number of beneficiaries starting from the divisor, while in the present case the solution is precisely their direct ratio; we can express it using the formula F3:

$$\sqrt{Q} = \frac{aD+bC}{C+D}$$

in which Q is the number of which the square root must be calculated, $C = Q - a^2$, $D = b^2 - Q$ in accordance with the numerical calculations performed above. Evidently, F3 is not identical to F2 or F1, but their respective numerators as 'dividend' and 'divisor' are part of it.

At first glance, it is not clear why a similar method already used for problems of a completely different kind was applied to calculate a square root, and why it is so effective. In fact, we will see examples in which its application will give good results. The second question can be answered more easily; in fact, if we substitute C and D by their respective expressions as a function of Q , a and b , we get

$$\begin{aligned} \frac{a(b^2-Q)+b(Q-a^2)}{Q-a^2+b^2-Q} &= \frac{ab(b-a)+Q(b-a)}{b^2-a^2} = \\ &= \frac{ab}{a+b} + \frac{Q}{a+b} \quad (\text{F4}) \end{aligned}$$

This expression can be considered the arithmetic mean between the harmonic mean of two approximations, one for defect (a) and the other for excess (b) and the ratio between the radicand Q and the arithmetic mean of a and b . In fact, F4 is equivalent to

$$\frac{(\frac{2ab}{a+b})+(\frac{2Q}{a+b})}{2}$$

which we can interpret as a correction of the harmonic mean between the defect and excess approximation.

I am unable to explain to myself how ancient Chinese mathematicians arrived at the F3 formula, since the considerations just made can justify why the method works, but nothing more. It seems to have to be ruled out that they started from F4, since switching to F3 is a complication. Perhaps, since calculating a square root implies quantities both greater and less than the radicand just as happens in problems solvable with analogous procedures, they tried to apply them to calculating the square root and found that the method, although not logically justified (at least according to modern criteria), was nevertheless useful and provided sufficiently accurate results.

EXAMPLES OF CALCULATING SQUARE ROOTS

Let's see some examples of approximate calculation of square roots of ten numbers using F4 and a computer, with the relative accuracy (ratio between the difference between the result given by F4 and that of a computer with ten exact digits, and the latter).

N	a ; b	F4	computer	accuracy
10	3 ; 4	3.142857142	3.162277660	0.006
50	7 ; 8	7.066666666	7.071067811	0.0006
200	14 ; 15	14.13793103	14.14213562	0.0003
300	17 ; 18	17.31428571	17.32050807	0.00035
500	22 ; 23	22.35555555	22.36067977	0.0002
1000	31 ; 32	31.61904761	31.62277660	0.0001
5000	70 ; 71	70.70921985	70.71067811	$2 \cdot 10^{-5}$
20000	141 ; 142	141.4204946	141.4213562	$6 \cdot 10^{-6}$
50000	223 ; 224	223.6062639	223.6067977	$2.4 \cdot 10^{-6}$
100000	316 ; 317	316.2274881	316.2277660	$1 \cdot 10^{-6}$

As you can see, as the radicand N increases, the approximation obtained from F4 seems to converge towards the one provided by the computer.

Any part of this article can be freely distributed and reproduced, as long as the author (E.F.Scriptor) is cited.

Turin, 19 December 2025